

*Appl. Statist.* (2018)

# On the ranking of test match batsmen

Richard J. Boys

*Newcastle University, Newcastle upon Tyne, UK*

and Peter M. Philipson

*Northumbria University, Newcastle upon Tyne, UK*

[Received September 2016. Revised June 2018]

**Summary.** Ranking sportsmen whose careers took place in different eras is often a contentious issue and the topic of much debate. We focus on cricket and examine what conclusions may be drawn about the ranking of test batsmen by using data on batting scores from the first test in 1877 onwards. The overlapping nature of playing careers is exploited to form a bridge from past to present so that all players can be compared simultaneously, rather than just relative to their contemporaries. The natural variation in runs scored by a batsman is modelled by an additive log-linear model with year, age and cricket-specific components used to extract the innate ability of an individual cricketer. Incomplete innings are handled via censoring and a zero-inflated component is incorporated in the model to allow for an excess of frailty at the start of an innings. The innings-by-innings variation of runs scored by each batsman leads to uncertainty in their ranking position. A Bayesian approach is used to fit the model and realizations from the posterior distribution are obtained by deploying a Markov chain Monte Carlo algorithm. Posterior summaries of innate player ability are then used to assess uncertainty in ranking position and this is contrasted with rankings determined via the posterior mean runs scored. Posterior predictive checks show that the model provides a reasonably accurate description of runs scored.

**Keywords:** Censoring; Overdispersion; Poisson random effects; Zero inflation

## 1. Introduction

There is much discussion in many sports, from the experts through to the fans, about who is the ‘greatest’. Discussions often conclude with the notion that it is impossible to obtain definitive answers. In many cases the game played out in the modern day, in front of the massed media with large teams of supporting staff dedicated to nutrition, fitness and psychology, bears little or no relation to the backdrop at the genesis of the sport. The richness of data that are now available, however, suggests that there may be merit in a sophisticated statistical approach to the problem.

The analysis of sports data has undergone a boom in recent years with statisticians and data analysts at the forefront. In baseball, for example, ‘sabermetrics’ has become an accepted term for the use of in-game statistical analysis (Marchi and Albert, 2013), and there is an increasing trend for sports science and data analysis being routinely performed by major sports organizations across the globe.

[Copyright line corrected after first online publication on 5th October 2018]

*Address for correspondence:* Richard J. Boys, School of Mathematics and Statistics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.  
E-mail: [richard.boys@newcastle.ac.uk](mailto:richard.boys@newcastle.ac.uk)

© 2018 The Authors Journal of the Royal Statistical Society: Series C (Applied Statistics) 0035–9254/18/68000  
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

In this paper we focus on the sport of cricket and look at the performance of test match batsmen. Cricket is a bat-and-ball game played between two teams of 11 players each on a cricket field, at the centre of which is a rectangular 22-yard-long pitch with a target called the wicket (a set of three wooden stumps topped by two bails) at each end. Each phase of play is called an innings during which one team bats, attempting to score as many runs as possible, while their opponents field. In test matches the teams have two innings apiece and, when the first innings ends, the teams swap roles for the next innings. This sequence can only be altered by the team batting second being made to 'follow-on' after scoring significantly fewer runs than the team batting first. Except in matches which result in a draw, the winning team is the team that scores the most runs, including any extras gained. Individual players start their innings with 0 runs and accumulate runs as play progresses, leading to a final score which is a non-negative count. The highest individual score in test cricket is 400 runs and the average score is around 30 runs. Smaller scores are more likely than larger scores as the aim of the opposition is to bowl out each batsman as quickly as possible and at the cost of as few runs as possible.

The earliest work on the statistical modelling of cricket scores was undertaken by Elderton (1945) and Wood (1945) who considered modelling samples of individual first-class cricket scores from both test matches and the County Championship (the domestic first-class cricket competition in England and Wales, sitting one level below test cricket) as a geometric progression and found evidence of a reasonable fit, although Wood commented that the 'series show discrepancies at each end, and particularly at the commencement' due to a larger-than-expected number of scores of 0, or 'ducks' in cricketering parlance. Incomplete ('not-out') scores were assumed to continue at the start of the next innings in Elderton (1945) (acknowledged as a 'pleasant fiction' by him) and treated as complete innings by Wood (1945). Later Pollard *et al.* (1977) investigated the distribution of runs scored by teams in county cricket and found that the negative binomial distribution offered a good fit. Scarf *et al.* (2011) confirmed this finding for runs scored in both batting partnerships and team innings in test cricket.

Kimber and Hansford (1993) considered the merits of the geometric distribution for samples of individual cricket scores from test and first-class matches, including Australia's domestic Sheffield Shield competition, along with 1-day internationals, concluding that 'there was little evidence against the... model in the upper tail' but rejecting its validity for low scores, mainly because of the excess of ducks in the data. Their work focused on an alternative batting average measure using a non-parametric product limit estimation approach. Some of these points will be revisited later. They also looked at the independence of cricket scores for a batsman and found 'no major evidence of autocorrelation' via a point process approach, surmising that 'it is quite reasonable... to treat scores as if they were independent and identically distributed observations'. Durbach and Thiart (2007) later concluded that batting scores can be considered to come from a random sequence on the basis of a study of 16 test match batsmen. We note that studies in other sports of a lack of independence of points or run scoring, sometimes referred to as the 'hot hand', have largely concluded that there is little evidence to support the notion of 'form' (Gilovich *et al.*, 1985; Tversky and Gilovich, 1989).

Published work in sports statistics covers a wide range of sports. Initially much of this work centred on the analysis of football and baseball, and focused on predicting future outcomes but now increasingly looks at gains that might be made by using an optimal strategy. The most famous model-based method that is used in cricket today is, of course, the Duckworth–Lewis–Stern formula (Duckworth and Lewis, 1998, 2004; Stern, 2009) for interrupted 1-day cricket matches, with subsequent modification by, for example, McHale and Asif (2013). Other work such as Silva *et al.* (2015) has looked at the effect of powerplay in such matches. In this paper the focus is instead on comparing past and current players, which is an area where relatively

little research has been done (Rohde, 2011; Radicchi, 2011; Baker and McHale, 2014), and we study test cricketers in particular. The innate ability of each player is modelled by taking into account the heterogeneous effect of aging on sporting performance, any year effects which act as a surrogate for changes to the game that may have made it easier or more difficult in certain eras, home advantage and some cricket-specific components. Berry *et al.* (1999) considered how to compare players from different eras in three, predominantly US-based, sports: baseball, golf and ice hockey. Their argument, which is adopted here, is that comparisons between modern day players and players from bygone eras are possible by considering the overlap in playing careers: modern players at the start of their careers will have played against older players at the end of their careers, which started much earlier, and these older players would, in their youth, have played against players from earlier eras once more. In such a way a bridge from the present to the past is formed.

The paper is structured as follows. The data are described in Section 2. The model description in Section 3 begins by outlining an initial model before introducing modifications to handle some nuances of cricket batting data. Sections 4 and 5 detail the prior and posterior distributions respectively along with the Markov chain Monte Carlo (MCMC) algorithm. Section 6 describes some of the results such as the posterior mean of player ability, and a ranking by this measure, and summaries of the posterior distribution of player rankings. The paper concludes with some discussion and avenues for future work in Section 7.

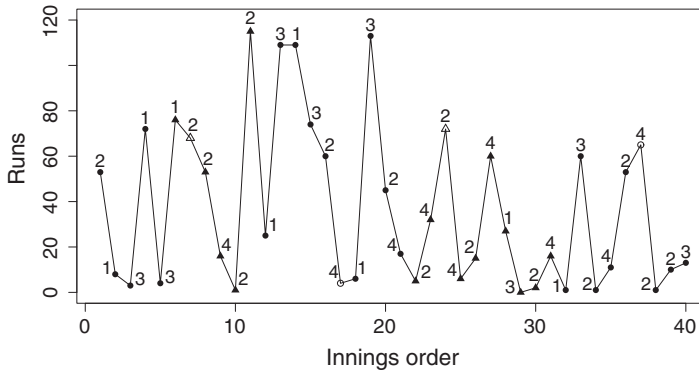
## 2. The data

Cricket is a highly data-driven sport, perhaps more than any other with the exception of baseball. Players' entire careers are typically judged by a one-number summary: their average. A large amount of data, typically in the form of scorecards, is available for all formats of cricket at international, domestic and even regional level. For some players even ball-by-ball data have been recorded (the Association of Cricket Statisticians and Historians have these data for Sir Jack Hobbs) although such a level of granularity is not generally available and so is not considered further here.

The data that are used in this paper consist of individual innings by all test match cricketers ( $n = 2855$ ) from the first test played in 1877 up to test 2269, in August 2017. There are currently 10 test playing countries and many more test matches are played today than at the time of the first test. Indeed for the first 12 years the combatants were exclusively England and Australia. A demonstration of the growth of test match cricket is given in Table 1. We note that, in contrast with the standard presentation of historical batting averages such as at <http://stats.espncricinfo.com/ci/content/records/282910.html>, we

**Table 1.** Timeline of test match milestones

<i>Test match</i>	<i>Year</i>	<i>Elapsed years</i>
1	1877	
100	1908	31
500	1960	52
1000	1984	24
2000	2011	27



**Fig. 1.** Innings played by Ian Bell against Australia between 2009 and 2015:  $\blacktriangle$ ,  $\triangle$ , away matches;  $\Delta$ ,  $\circ$ , not-out innings; 1–4, innings number

include all batsmen irrespective of the number of career innings played. However, in keeping with other lists, we do not include World Series Cricket matches as these matches are not considered official test matches by the International Cricket Council (ICC).

In addition to runs scored, the data contain other useful information such as the venue, the opposing team and whether or not the batsman's innings is incomplete (this can happen for a variety of reasons; see Section 3.1.1). Thus we can determine whether a test match is played at 'home' and investigate the extent of any home advantage. Note that 29 tests have been played at neutral venues and we class these as away matches for both teams. The data also include the match innings index, which is potentially important as (generally) the conditions in the final innings of a test match are at their worst for batting and the pressure, due to the game situation, is often at its highest—it is an axiom of cricket that batting last is difficult.

Some aspects of the game that have changed over time are not explicitly recorded in the data: at one stage tests were 'timeless', continuing until a result was achieved; the number of balls in an over has varied between 4, 5, 6 and 8; pitches were uncovered and left exposed to the elements up to around 1960; the introduction of limited overs international cricket in 1971 along with the recent advent of Twenty20 cricket in 2003 and the abolition of 'amateur' status in 1962. Together these aspects may have affected the performance of batsmen, particularly possible changes to pitches and changes to the game dynamic that are induced by the shorter formats. We shall consider these on annual and decade scales respectively.

A typical profile of batting scores is given in Fig. 1. This plot shows the runs that were scored in the test match innings of England batsman Ian Bell against Australia between 2009 and 2017. The innings are shown in sequential order and away matches and not-out innings are indicated by the plotting symbol. Note that, although Bell was in the England side throughout this period, he did not bat in many innings. This feature is typical and can be due to many factors such as big wins where the follow-on was deployed and the winning team did not need to bat a second time, or if the match is drawn because of bad weather or running out of time. The figure highlights the capricious nature of batting and suggests that, whereas year, aging and game-specific effects may affect run scoring on an overall level, the innings-by-innings variation is considerable.

### 3. The model

Runs scored in an innings are counts and a natural starting point is to consider modelling them via the Poisson distribution, with

$$X_{ijk} | \lambda_{ijk} \stackrel{\text{indep}}{\sim} \text{Po}(\lambda_{ijk}), \quad i = 1, \dots, 2855, \quad j = 1, \dots, n_i, \quad k = 1, \dots, n_{ij},$$

where  $i$  is the player index,  $j$  is the year index and  $k$  is the innings index so  $X_{ijk}$  represents the number of runs scored by player  $i$  in his  $j$ th year during his  $k$ th innings of that year. Also  $n_i$  and  $n_{ij}$  denote respectively the number of years in the career of player  $i$  and the number of innings played during year  $j$  in the career of player  $i$ .

Notation for other available information is as follows. For player  $i$  in the  $k$ th innings of their  $j$ th career year,  $y_{ijk}$  is the year in which the innings was played,  $a_{ijk}$  is the age of the player,  $h_{ijk}$  indicates whether the innings was played in the batsmen's home country (1, home; 2, away),  $m_{ijk}$  is a within-match innings index (which is different from the within-year innings label  $k$ ),  $o_{ijk}$  is the opposition's country and  $e_{ijk}$  is an indicator for the era of play, which here is considered on a decade scale. These last two pieces of information together allow us to study possible changes to the performance of a country over time.

Within this Poisson framework we adopt a log-linear model for the run scoring rate which includes the main components that are thought to influence its outcome, with

$$\log(\lambda_{ijk}) = \theta_i + \delta_{y_{ijk}} + f_i(a_{ijk}) + \zeta_{h_{ijk}} + \nu_{m_{ijk}} + \xi_{o_{ijk}} + \omega_{o_{ijk}, e_{ijk}} \quad (1)$$

where  $\theta_i$  represents the ability of player  $i$ , the difficulty of the year is captured through  $\delta_l$  (the data span 141 years) and  $f_i(a)$  is a player-specific aging function, of which more in a moment. The remaining terms in the model are game specific, representing respectively the effect of playing at or away from home, the match innings effects, the quality of the opposition and an interaction term allowing for the quality of the opposition to change over different eras. Here we take eras to be decades to reduce the number of parameters in the model.

The player ability parameter captures the contribution to runs scored that can be attributed to the fundamental talent of the player. As mentioned earlier, aging can have a strong effect on sporting performance so we incorporate an individual quadratic aging function as suggested by Albert's discussion in Berry *et al.* (1999), namely

$$f_i(a) = -\alpha_{2i}(a - \alpha_{1i})^2,$$

where  $\alpha_{1i}$  is the age at which the peak is attained and  $\alpha_{2i}$  is the curvature of the function which measures the rate at which the individual matures and declines.

The year effects are a composite of several factors: clear-cut changes such as depth of competition (more test playing countries), game focus (scoring rates are far higher in modern times and there are fewer draws) and law changes (e.g. fewer bouncers per over allowed to make batting easier) whereas others are more subtle, for instance technological advances and game conditions (most pitches are prepared to last 5 days to ensure maximum profit). We anticipate that the year effects vary smoothly over time and allow for this by using a random-walk prior; see Section 4. The year effects also need to be standardized for identifiability and so we compare these effects relative to the final year in the data set (2017) by taking  $\delta_{141} = 0$ .

The remaining terms in the model account for home advantage, which is common in many team sports, and two further context-specific effects to represent that, as pitches deteriorate, and the match situation becomes more acute, batting may become more difficult and to take into account the quality of the opposition. We set the home effect as the reference level (by taking  $\zeta_1 = 0$ ) and measure the effect of playing away by  $\zeta_2$ . The innings effects are represented through  $\nu_g$  to reflect the difficulty of innings  $g$  where  $g = 1, 2, 3, 4$  is the innings of the match in which the runs were scored, and with  $\nu_1 = 0$  for identifiability. The quality of the opposition is taken into account via  $\xi_q$  for  $q = 1, \dots, 10$  to represent the 10 test playing countries, some of

which have traditionally been stronger than others. Here we number the countries alphabetically. For identifiability we shall take Australia, the first team in the alphabetical ordering of the test match playing nations, as the reference opposition country, with  $\xi_1 = 0$ . Further, the opposition–decade interactions are compared with the final decade (by taking  $\omega_{1:10,14} = 0$ ) and with that of Australia (by also taking  $\omega_{1,1:13} = 0$ ).

Thus in this model  $\exp(\theta_i)$  is the average number of runs per innings scored by player  $i$  when he is at his peak age, playing at home against Australia, and in the first innings of a test match taking place in 2017.

### 3.1. Poisson random-effects model

There is substantial variation in individual innings-by-innings cricket scores. As such, the inherent assumption of equidispersion in the Poisson model is unlikely to hold. Under this model and considering players who score on average 10 or more runs per innings we would expect their distribution of scores to be broadly Gaussian. However any follower of cricket would intuitively feel that this is not so and that excess variability to that provided by the Poisson model is present. The data in Fig. 1 on Ian Bell are typical of many other players and show extra-Poisson variation with censored observations and perhaps more ducks than expected. We now augment the model to allow for each of these features.

We allow for the extra-Poisson variation by introducing random effects, acting multiplicatively on the Poisson mean parameter, so that

$$X_{ijk} | \lambda_{ijk}, v_{ijk} \stackrel{\text{indep}}{\sim} \text{Po}(\lambda_{ijk} v_{ijk}).$$

There are many possible choices of distribution for the random effects  $v_{ijk}$ , such as gamma, log-normal, inverse Gaussian or general power transforms (Hougaard *et al.*, 1997). We shall use the gamma distribution as this gives a negative binomial distribution for the number of runs after integrating over  $v_{ijk}$  (Cameron and Trivedi, 1986; Greene, 2008). This choice allows a direct comparison with earlier work, particularly as the geometric distribution is a special case. For further flexibility we allow the random-effects distributions to be player specific, reflecting that player characteristics, such as aggression, can lead to substantial differences in variability between players of comparable ability. Thus we take  $v_{ijk} \sim \text{Ga}(\eta_i, \eta_i)$ , with  $E(v_{ijk}) = 1$  and  $\text{var}(v_{ijk}) = 1/\eta_i$ . Therefore (marginally) we use a negative binomial model for runs scored, with

$$X_{ijk} | \lambda_{ijk}, \eta_i \stackrel{\text{indep}}{\sim} \text{NB}\{\eta_i, \eta_i/(\eta_i + \lambda_{ijk})\}.$$

Introducing the random effects makes no change to the (marginal) mean but has inflated the (marginal) variance to  $\text{var}(X_{ijk}) = \lambda_{ijk}(1 + \lambda_{ijk}/\eta_i)$ , with the basic Poisson model being recovered as  $\eta_i \rightarrow \infty$ . This form of variance function is appropriate for modelling batting scores as the variability is smaller for players of lesser ability (they have a more restricted range of runs scored and rarely achieve high numbers of runs) and larger for players of higher ability (although they score high numbers of runs, they will typically also have innings with very low scores). The form of the negative binomial success probability is cumbersome and so to simplify the exposition we shall use  $\beta_{ijk} = \lambda_{ijk}/\eta_i$ .

We now augment the model to deal with

- (a) incomplete scores—innings where the batsmen is not dismissed—and
- (b) potential zero inflation in the data—more ducks (0 scores) in the data than the model suggests.

### 3.1.1. Censoring

Approximately 13% of innings are incomplete, which is referred to as ‘not out’ in cricketing vernacular, typically because of the completion of a team’s innings, which, by necessity, must include one incomplete innings at the fall of the final wicket, or two incomplete innings in a successful run chase (or if the match has not been completed because of adverse weather or running out of time). Incomplete innings can also happen when the team captain ‘declares’ and brings the innings to a premature close (typically to aid the prospect of victory) and this can result in either one or two incomplete innings. Historically, cricket has dealt with incomplete innings in a somewhat *ad hoc* manner whereby the runs are added to the numerator in the batting average without any increment to the denominator. Clearly such innings ought not to be dealt with in the same way as a complete innings and the standard cricketing treatment can exaggerate the contribution of incomplete innings and thereby affect the batting average. From a statistical viewpoint, a not-out innings is simply a censored observation. Kimber and Hansford (1993) claimed that ‘ $x$  not-out is representative of all scores of  $x$  or more’ and so we assume non-informative censoring. Thus, denoting a not-out (censored) innings by the binary variable  $c$ , the likelihood contribution from player  $i$ , for the  $k$ th innings of the  $j$ th year of his career, is

$$\left\{ \binom{x_{ijk} + \eta_i - 1}{x_{ijk}} \beta_{ijk}^{x_{ijk}} / (1 + \beta_{ijk})^{\eta_i + x_{ijk}} \right\}^{1 - c_{ijk}} P(X_{ijk} \geq x_{ijk})^{c_{ijk}}, \quad (2)$$

where  $X_{ijk}$  has a negative binomial distribution.

### 3.1.2. Zero inflation

After ignoring censored 0s, ducks account for almost 11% of the observations in the data. Even Sir Donald Bradman (with a test batting average of 99.94) had a modal score of 0 with seven ducks out of 80 innings. This high proportion of 0s is likely to be due to players being vulnerable early in their innings (Brewer, 2008), taking time to acclimatize to conditions and ‘get their eye in’ rather than to some other process that causes scores to be necessarily zero. Thus the proportion of ducks is likely to be higher than expected by using the Poisson random-effects model and so we modify the model to allow for this inflation of 0s. We also allow for different levels of zero inflation for each player.

There are two basic ways of dealing with zero inflation. One way is to model the probability of obtaining a 0 by a mixture of the primary model and a point mass at zero (Lambert, 1992) and the other is to use a hurdle model which contains a model for 0 counts (the hurdle component) and a separate model for the strictly positive counts (once the hurdle, a batsman playing a scoring stroke for instance, has been cleared). Hurdle models are particularly popular in the economics literature; see, for example, Gurmu (1997, 1998). They are the natural choice when the 0s are entirely structural, such as in a biological process (Ridout *et al.*, 2001) or a weather pattern (Scheel *et al.*, 2013). We favour the mixture representation as this can be interpreted as the number of ducks being a mixture of (the Poisson random-effects) model-based 0s and a component representing the increased vulnerability of a batsman early in an innings. This representation has the additional advantage (which is not followed up here) of providing a framework for generalizing the model to inflate other scores, such as 4 or 6, that may occur more frequently because they are achievable with a single scoring stroke, i.e. via a ‘four’ or a ‘six’.

The excess 0s are assumed to be unrelated to the other effects and so we model the probability of obtaining a (completed) duck for player  $i$  as

$$P(X_{ijk} = 0) = \pi_i + (1 - \pi_i) / (1 + \beta_{ijk})^{\eta_i}.$$

As the player-specific parameter  $\pi_i \rightarrow 0$ , the zero-inflated component diminishes and the number of (completed) ducks is well described by an orthodox Poisson random-effects model. Thus, denoting a batsman with a (completed) duck by the binary variable  $d$ , the likelihood contribution from player  $i$ , for the  $k$ th innings of the  $j$ th year of his career, is amended from that in equation (2) to

$$\begin{aligned} & \{\pi_i + (1 - \pi_i)/(1 + \beta_{ijk})^{\eta_i}\}^{(1-c_{ijk})d_{ijk}} \\ & \times \left[ (1 - \pi_i) \left\{ \binom{x_{ijk} + \eta_i - 1}{x_{ijk}} \beta_{ijk}^{x_{ijk}} / (1 + \beta_{ijk})^{\eta_i + x_{ijk}} \right\}^{1-c_{ijk}} P(X \geq x_{ijk})^{c_{ijk}} \right]^{1-d_{ijk}}. \end{aligned} \quad (3)$$

Introducing a zero-inflation effect also reduces the expected number of runs scored by a factor of  $1 - \pi_i$ .

#### 4. The prior distribution

We need to construct a joint prior distribution for the many parameters in this model. In general, we have chosen to describe our prior beliefs by taking fairly weak independent priors for each parameter component. This has the benefit of ‘letting the data speak’ and gives our results a reasonable level of robustness against our choice of prior.

We adopt a random-effects style (or hierarchical) prior for the player-specific ability parameters in which ability varies between batsmen by taking

$$\theta_i | \mu_\theta, \sigma_\theta \stackrel{\text{indep}}{\sim} N(\mu_\theta, \sigma_\theta^2).$$

We also take semiconjugate prior distributions for the ability parameters, with  $\mu_\theta \sim N(m_\mu, s_\mu^2)$  and  $\sigma_\theta^2 \sim \text{IG}(a_\sigma, b_\sigma)$ , where  $\text{IG}(a, b)$  denotes the inverse gamma distribution with mean  $b/(a - 1)$ . It was felt that the median number of runs scored across all innings (including not-out innings) would be around 20 and so we take  $m_\mu = \log(20)$ . Also the variability between decades of runs scored was likely to be within a 60%-fold increase or decrease and so we take  $s_\mu = 0.25$  (as  $\exp(0.5) \simeq 1.6$ ). Variation of player ability was thought to be typically about a fourfold increase or decrease around the decade mean, giving  $\sigma_\theta^2$  a mean of around 0.5, and that the probability that this fold increase or decrease would exceed 10 was around 5%. Together these requirements give a prior distribution with (roughly)  $a_\sigma = 3$  and  $b_\sigma = 1$ .

It was felt that the year effects  $\delta_l$  should vary fairly smoothly in time and that prior beliefs were less certain for years going increasingly further into the past. Therefore, together with the identifiability constraint  $\delta_{141} = 0$ , we use the (backward) simple random walk  $\delta_l = \delta_{l+1} + \sigma_\delta \epsilon_l$ ,  $l = 140, \dots, 1$ , where the  $\epsilon_l$  are independent standard normal quantities. To see its smoothing role, it is useful to note that this random walk induces

$$\delta_l | \delta_{(l)}, \sigma_\delta \sim N\left(\frac{\delta_{l-1} + \delta_{l+1}}{2}, \frac{\sigma_\delta^2}{2}\right), \quad \text{for } l = 2, \dots, 140,$$

with  $\delta_1 | \delta_{(1)}, \sigma_\delta \sim N(\delta_2, \sigma_\delta^2)$ , where  $\delta_{(l)} = (\delta_i, i \neq l)$  represents all of the year effects except year  $l$ . For notational convenience we write  $\delta$  for the year effects  $\delta_{(141)}$ . These descriptions lead to the prior distribution of the year effects being  $\delta | \sigma_\delta \sim N_{140}(\mathbf{0}, \sigma_\delta^2 Q^{-1})$  where the inverse correlation matrix  $Q$  has the tridiagonal structure



$$Q = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

The parameter  $\sigma_\delta$  describes the smoothness of the year effects and, as this impacts player ability on an exponential scale, it was felt that  $\sigma_\delta^2$  should have an  $\text{IG}(a_\delta, b_\delta)$  prior distribution with mean 0.01 and only a 5% probability of exceeding 0.03. This leads (roughly) to a choice of prior parameters  $a_\delta = 2$  and  $b_\delta = 0.01$ .

We now consider the prior distributions for the remaining parameters, beginning with the game-specific parameters: the effect of playing away  $\zeta_2$ , the innings effects  $\nu_{2:4}$ , the quality of the opposition  $\xi_{2:10}$  and the opposition–era interactions  $\omega_{2:10,1:13}$  (recall that  $\zeta_1 = \nu_1 = \xi_1 = \omega_{1:10,14} = \omega_{1,1:13} = 0$  for identifiability). The strength of our opinion on their potential size is quite weak and so we give these parameters zero-mean normal prior distributions with standard deviation 0.5, this taken to equate to a 95% prior credible interval for these effects spanning an increase or decrease of around 2.7 fold on the runs scored. Our prior beliefs about the player-specific aging function are that the peak age is around 30 years old and that the rate of maturity and decline of players at 7 years respectively before and after their peak is respectively roughly  $\frac{2}{3}$  and  $-\frac{2}{3}$ . We represent our fairly weak prior beliefs by taking  $\alpha_{1i} \sim N(30, 4)$  and  $\alpha_{2i} \sim \text{LN}(-3, 9)$ .

Previous studies have considered a geometric random-effects distribution for runs scored and so we give the individual random-effects heterogeneity parameters  $\eta_i$  a log-normal prior with unit prior median, but we also make this prior fairly weak by taking  $\eta_i \sim \text{LN}(0, 1)$ . Our prior beliefs about the individual zero-inflation parameters  $\pi_i$  are captured by a  $\text{beta}(a_\pi, b_\pi)$  distribution with mean 0.1 and only a 5% probability of  $\pi_i$  exceeding 0.3. This leads (roughly) to a choice of prior parameters  $a_\pi = 1$  and  $b_\pi = 9$ .

## 5. The posterior distribution

The posterior density can be factorized as

$$\pi(\boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\pi} | \mathbf{x}, \mathbf{c}, \mathbf{d}) \propto \pi(\mathbf{x}, \mathbf{c}, \mathbf{d} | \boldsymbol{\kappa}, \boldsymbol{\eta}, \boldsymbol{\pi}) \pi(\boldsymbol{\kappa}) \pi(\boldsymbol{\eta}) \pi(\boldsymbol{\pi})$$

with  $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\kappa})$ , where  $\mathbf{x}$ ,  $\mathbf{c}$  and  $\mathbf{d}$  are the vectors of runs scored and associated censoring and duck indicators respectively, and  $\boldsymbol{\kappa} = (\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma_\delta, \boldsymbol{\alpha}, \zeta_2, \boldsymbol{\nu}, \boldsymbol{\xi}, \boldsymbol{\omega})$  contains the remaining parameters in the model, with  $\boldsymbol{\nu} = (\nu_{2:4})$ ,  $\boldsymbol{\xi} = (\xi_{2:10})$  and  $\boldsymbol{\omega} = (\omega_{2:10,1:13})$ . This posterior distribution is analytically intractable and we therefore turn to a sampling-based approach and make inferences via the use of MCMC methods.

In our MCMC scheme we generally use Metropolis–Hastings steps with symmetric normal random-walk proposals on an appropriate scale and centred on the current value, e.g. on the log-scale for positive quantities or the logit scale for quantities restricted to  $(0, 1)$ . Overall we have found that this strategy works well except for updates to the year effects  $\boldsymbol{\delta}$ . Here Gibbs updates are available for each component  $\delta_l$  but their full conditional distributions depend strongly on the values taken by the year effects on either side, i.e.  $\pi(\delta_1 | \cdot) = \pi(\delta_1 | \delta_2, \cdot)$  and  $\pi(\delta_l | \cdot) = \pi(\delta_l | \delta_{l-1}, \delta_{l+1}, \cdot)$ ,  $l \neq 1$ . This is not surprising given the dependence structure in the prior distribution for  $\boldsymbol{\delta}$ . It is well known that such strong dependence can lead to poor mixing such as that in the distribution of hidden states in hidden Markov models. Also this strong dependence prohibits the use of software such as JAGS (Plummer, 2004) to obtain posterior realizations in

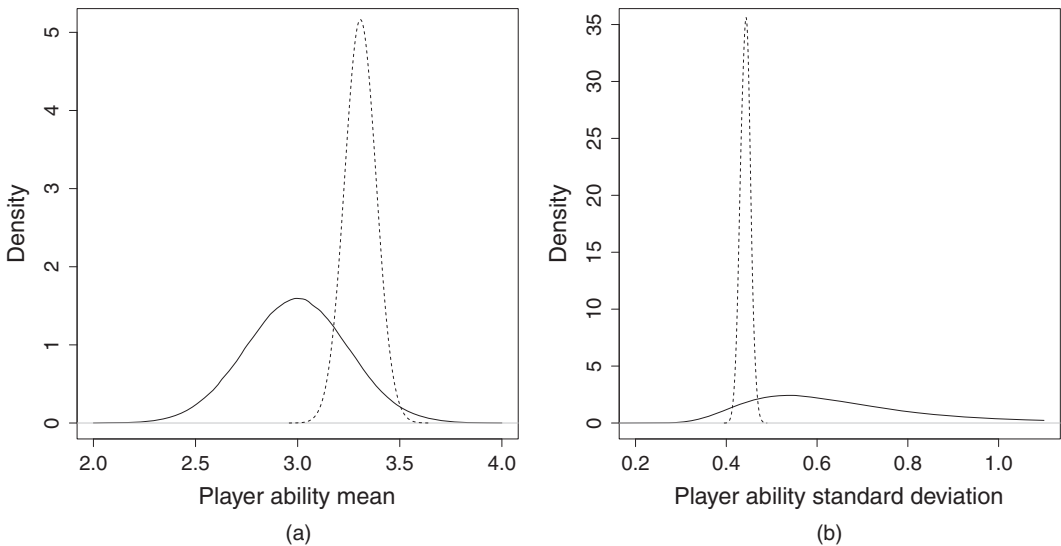
a timely manner. Instead we follow Gamerman (1997) and construct a normal proposal distribution for  $\delta$  via a Taylor series approximation to its full conditional distribution; see the on-line supplementary materials for further details. We have found that this strategy greatly improves the mixing of the scheme.

## 6. Results

An implementation of the MCMC scheme in R (R Core Team, 2014) is available from <https://github.com/petephippison/Ranking-Test-batsmen> together with the data. We report here results from a typical run of the MCMC scheme which used a burn-in of 5000 iterations and was then run for a further 200000 iterations, with this (converged) output thinned by taking every 20th iterate. This gave a posterior sample of  $N = 10000$  (almost) un-auto-correlated values for analysis. Convergence was assessed through a variety of graphical and numerical diagnostics via the R package *coda* (Plummer *et al.*, 2006).

### 6.1. Random-effects distribution for player ability

The (marginal) posterior distributions for the mean and standard deviation ( $\mu_\theta$  and  $\sigma_\theta$ ) of the random-effects distribution for player ability are shown in Fig. 2. Clearly the data have been quite informative. We can obtain a quick understanding of this posterior distribution by looking at its implication for the (random-effects) distribution of the number of runs scored (by players at their peak age, playing at home against Australia, and in the first innings of a test match taking place in 2017). Ignoring the (player-specific) zero-inflation effect, the five-number summary (minimum–lower quartile–median–upper quartile–maximum) for the median number of runs scored ( $\exp(\mu_\theta)$ ) is 24.7–26.4–27.3–28.3–30.2, and that for the average number of runs scored ( $\exp(\mu_\theta + \sigma_\theta^2/2)$ ) is 25.2–26.9–27.9–28.8–30.8. These distributions seem reasonable after taking into account that the zero-inflation parameters  $\pi$  are around 8% (see Section 6.4).



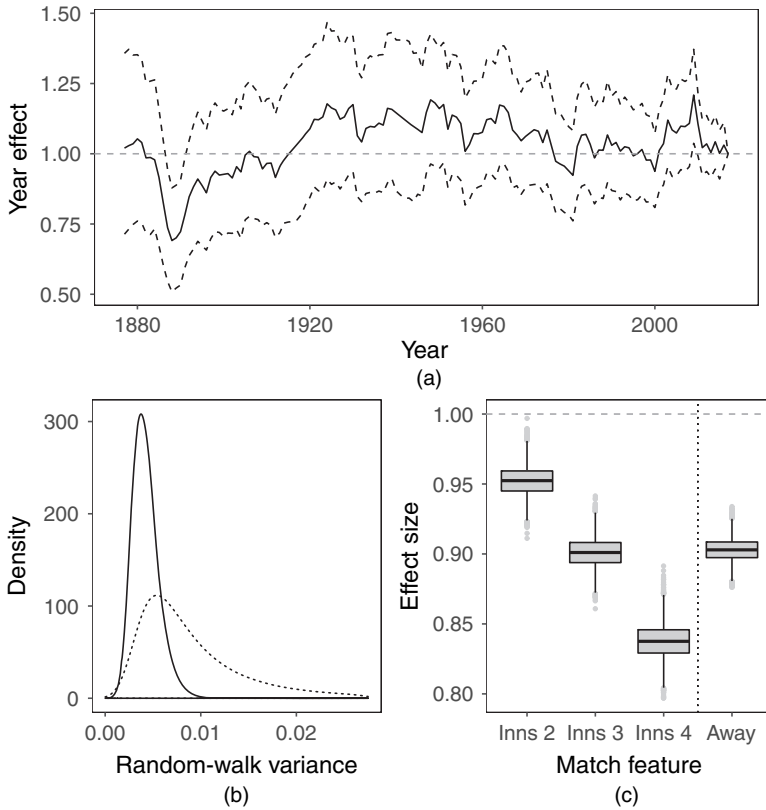
**Fig. 2.** Prior (—) and posterior (-----) density plots for ability parameters (a)  $\mu_\theta$  and (b)  $\sigma_\theta$

### 6.2. Year effects

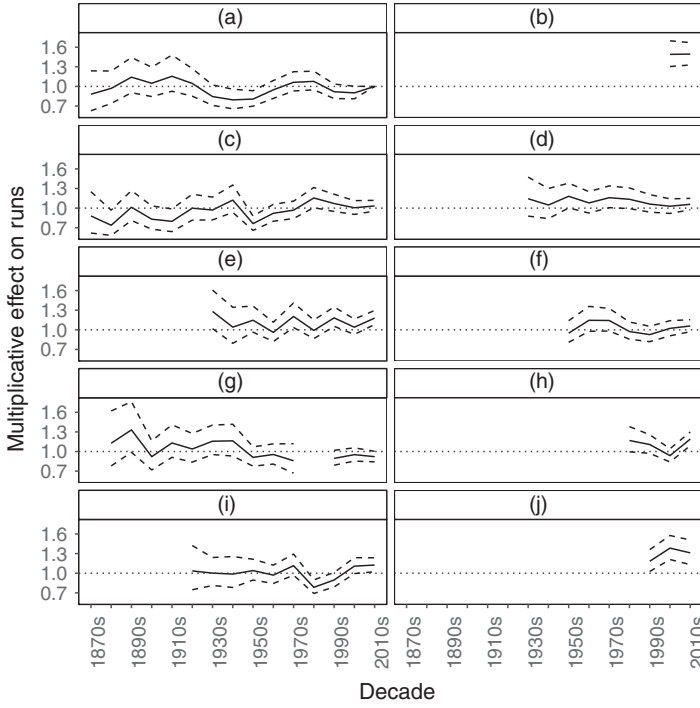
The posterior distribution for the year effects is summarized in Fig. 3(a). The effects are shown on an exponential scale and so represent the multiplicative effect on run scoring for each year, relative to playing against Australia in the most recent year (2017), here shown by the horizontal broken line. It is clear that there are very few important year effects, with the main (and negative) deviation being 1887–1891, which is a period when it is widely acknowledged that bowling conditions were favourable. The next strongest (and positive) deviation occurred in 2009, which is a year featuring four of the 16 highest team scores of all time. Fig. 3(b) shows the prior and posterior distribution of the smoothing parameter  $\sigma_\delta^2$  for the year effects. The slight shift in the posterior towards lower values suggests that the prior distribution did not oversmooth.

### 6.3. Home advantage, innings and opposition effects

Fig. 3(c) provides a visual comparison of the size of the multiplicative effect on run scoring when batting in different innings and playing away from home. Note that these effects are all relative to playing at home against Australia in the first innings in 2017, which is represented by the broken horizontal line. The effect of playing away from home on runs scored,  $\exp(\zeta_2)$ , has posterior mean 0.90 and 95% confidence interval (0.89, 0.92). Thus, there is a pronounced detrimental



**Fig. 3.** (a) Posterior mean and central 95% bands for the multiplicative year effects  $\exp(\delta_t)$ , (b) posterior (—) and prior (-----) density plots for the smoothing parameter  $\sigma_\delta^2$  and (c) boxplots of the posterior distributions for the multiplicative match innings effects  $\exp(\nu_m)$  and playing away effect  $\exp(\zeta_2)$



**Fig. 4.** Posterior mean and central 95% bands for the multiplicative opposition effects by decade,  $\exp(\xi_q + \omega_{qd})$ : (a) Australia; (b) Bangladesh; (c) England; (d) India; (e) New Zealand; (f) Pakistan; (g) South Africa; (h) Sri Lanka; (i) West Indies; (j) Zimbabwe

effect of playing away from home, leading to batsmen scoring 10% fewer runs. This finding is consistent with that found for ‘home advantage’ in other sports (Pollard and Pollard, 2005; Jones, 2007; Baio and Blangiardo, 2010). The posterior means (with 95% confidence intervals) for the second-, third- and fourth-innings effects ( $\exp(\nu_{2:4})$ ) are 0.95 (0.93, 0.97), 0.90 (0.88, 0.92) and 0.84 (0.81, 0.86) respectively, with the reference value of 1 for the first innings. These effects act multiplicatively on run scoring. Hence, performance decreases as the match goes on, with the innings effect at its strongest in the final innings of the match, as cricketing folklore would have predicted. The second innings of a test match is tougher than the first innings with a reduction of 5% in runs scored, but the effect increases to a 10% reduction in runs scored in the third innings and a 16% reduction in the final innings (compared with the first innings). It is interesting to see that the effect of batting in the third innings and that of playing away from home are very similar.

Fig. 4 displays the posterior means and associated 95% intervals of  $\exp(\xi_q + \omega_{qd})$  for the 10 test playing countries ( $q = 1, \dots, 10$ ) over the 15 decades ( $d = 1, \dots, 15$ ) during which test cricket has been played. As mentioned earlier, fewer countries played test cricket when it began as an international sport. The estimates in each case are relative to the strength of the current Australian test team (represented by the horizontal dotted line on each plot). There are 20 instances of opposition effects that show appreciable deviation from that of Australia in the most recent decade: these are split as six instances of an opposition being significantly more difficult to score runs against than the current Australia team and 14 cases where the opposition are easier to score runs against. The largest deviations (and with the lowest posterior means)

were England in the 1880s and 1950s, and the West Indies in the 1980s, each causing a 20–25% reduction in average runs scored.

The two newest test playing nations, Bangladesh and Zimbabwe, have struggled at times to be competitive and the three largest (significant) posterior means are for these two countries. Batsmen have preyed on this weakness, scoring on average over 50% more runs against Bangladesh and over 30% more runs against Zimbabwe. New Zealand were also relatively weak when they first played test cricket (in the 1930s) with batsmen scoring on average 30% more runs. Other noteworthy examples of weaker opposition were India in the 1950s, India and New Zealand in the 1970s and England in the 1980s. In each case batsmen scored on average around 20% more runs against these countries in these decades.

We investigated the sensitivity of our conclusions on opposition effects to using 5-year time periods rather than decades and found very little difference. Also there is no need to standardize opposition scores against the current Australia side and it is straightforward to standardize scores against any opposition team in any decade. We provide the results for any choice of team and decade via an RShiny application, which is available from <https://petephilipson.shinyapps.io/opposition/>.

#### 6.4. Random-effects heterogeneity and zero inflation

Five-number summaries of the posterior means and standard deviations for the player-specific random-effects heterogeneity parameters  $\eta_i$  are 0.48–0.87–1.01–1.16–2.19 and 0.07–0.24–0.40–0.54–0.89 respectively. The posterior distributions for a number of batsmen show a clear deviation from the geometric model ( $\eta_i = 1$ ) for cricket scores postulated by Elderton (1945) and Wood (1945). We note that they did not account for zero inflation (or censoring) but Wood did remark on a lack of fit for scores of 0.

Five-number summaries of the posterior means and standard deviations for the player-specific zero-inflation parameters  $\pi_i$  are 0.01–0.06–0.08–0.11–0.34 and 0.01–0.04–0.06–0.08–0.15 respectively. The posterior distributions show clearly both evidence for zero inflation in test match cricket scores and variation between players. The modal batsman's score in test cricket is 0, and the commonly held belief that batsmen are at their most vulnerable at the onset of their innings is a plausible explanation here. Posterior means of the  $\pi_i$  for the top 30 ranked batsmen are included in Table 2. The excess of 0s that was observed by Wood is a genuine feature of test cricket scores. It is interesting to note the discussion on the use of the standard cricket batting average summary in Kimber and Hansford (1993): they pointed out that such a measure is a consistent estimate only if the scores follow a geometric distribution.

#### 6.5. Individual aging

We determine the aging profile for a batsman by examining the posterior distribution of their expected runs scored at various ages  $a$ , i.e.  $(1 - \pi) \exp\{\theta + f(a)\}$ . Fig. 5 shows posterior mean profiles (and central 95% bands) for a selection of players of broadly similar ability but with quite different aging profiles. Also included in the plot are the posterior mean-adjusted runs scored for each player, i.e. the posterior mean of

$$\sum_{j,k:a_{ijk}=a} x_{ijk} \exp\{-(\delta_{y_{ijk}} + \zeta_{h_{ijk}} + \nu_{m_{ijk}} + \xi_{o_{ijk}} + \omega_{o_{ijk}, e_{ijk}})\} / n_{ia}$$

where  $n_{ia}$  is the number of completed innings played by player  $i$  at age  $a$ . Fig. 5 shows that the quadratic function largely captures the aging profiles, particularly when taking account of the

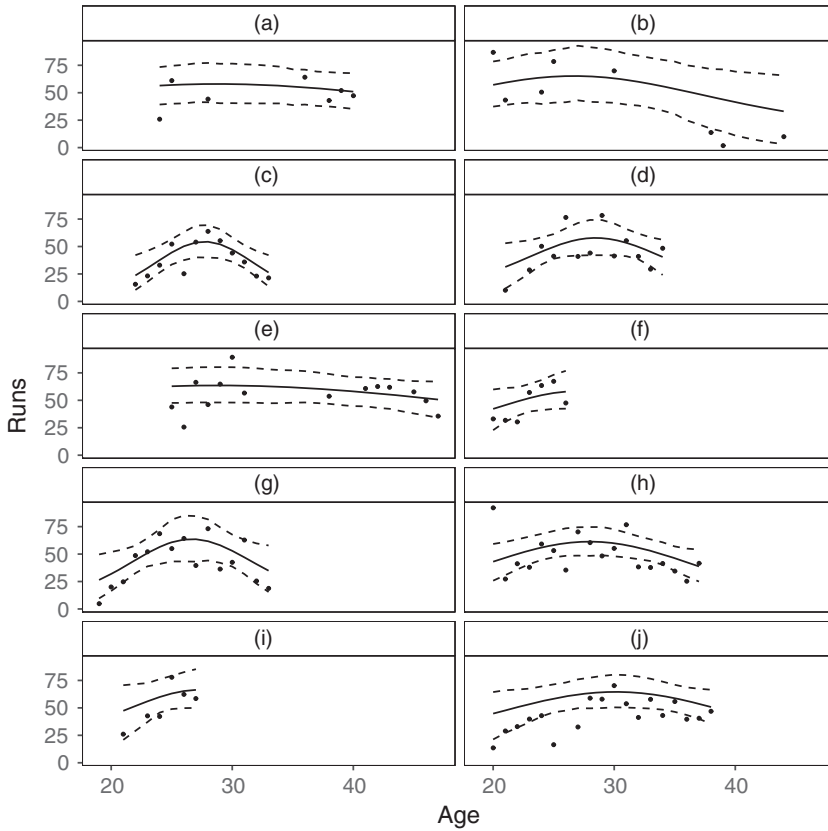
**Table 2.** Player rankings (ordered by posterior mean runs at peak age) together with posterior means for peak age and zero-inflation proportions, and summaries of player rank distributions

Rank	Name	Debut	Innings	Runs	Standard deviation	Peak age	Zero inflation (%)	Median rank (95% confidence interval)
1	D. G. Bradman	1928	80	93.7	14.3	28.2	7	2 (1–14)
2	S. P. D. Smith	2010	100	66.3	10.1	27.9	2	27 (3–158)
3	G. S. Sobers	1954	160	64.1	7.9	27.8	5	33 (5–137)
4	G. A. Headley	1930	40	63.2	12.8	27.0	4	40 (2–360)
5	C. L. Walcott	1948	74	63.2	9.6	28.4	2	38 (4–206)
6	S. R. Waugh	1985	260	62.0	7.7	29.6	5	43 (7–168)
7	H. Sutcliffe	1924	84	61.9	9.0	28.2	3	44 (5–218)
8	M. D. Crowe	1982	131	61.6	10.7	27.0	4	48 (4–304)
9	J. B. Hobbs	1908	102	61.4	8.2	28.4	4	45 (6–204)
10	J. H. Kallis	1995	280	61.3	6.8	28.6	3	47 (8–144)
11	S. R. Tendulkar	1989	329	61.2	6.4	27.2	2	46 (10–139)
12	E. D. Weekes	1948	81	60.5	8.6	27.8	5	52 (6–233)
13	R. T. Ponting	1995	285	59.8	6.8	27.8	3	55 (10–189)
14	W. R. Hammond	1927	140	59.4	7.4	28.1	2	59 (9–212)
15	R. G. Pollock	1963	41	59.2	11.7	27.7	3	69 (3–406)
16	K. F. Barrington	1955	131	58.9	7.1	28.6	2	62 (11–217)
17	L. Hutton	1937	138	58.7	7.3	28.1	2	65 (10–224)
18	B. C. Lara	1990	232	58.6	5.8	28.3	3	64 (16–173)
19	A. R. Border	1978	265	58.5	6.0	27.9	2	64 (14–180)
20	K. S. Williamson	2010	110	58.0	10.4	27.7	3	81 (4–343)
21	Y. Khan	2000	213	57.7	6.6	28.7	5	73 (13–219)
22	K. C. Sangakkara	2000	233	57.5	5.8	28.9	2	72.5 (16–198)
23	R. Dravid	1996	286	57.5	5.7	27.8	1	73 (17–198)
24	G. S. Chappell	1970	151	57.4	6.7	28.2	5	75 (12–242)
25	A. C. Voges	2015	31	57.2	14.2	28.5	4	91 (3–644)
26	H. M. Amla	2004	183	57.2	8.5	28.5	2	80 (9–329)
27	J. E. Root	2012	107	57.1	7.8	27.7	2	79 (9–303)
28	A. Flower	1992	112	56.9	7.3	28.5	2	80 (12–274)
29	S. M. Gavaskar	1971	214	56.7	6.1	28.0	2	81 (18–226)
30	M. Yousuf	1998	156	56.6	7.2	28.6	5	87 (11–267)

(posterior) uncertainty on the mean-adjusted scores (which are not shown). The posterior mean of the peak ages,  $\alpha_{1i}$ , is typically late 20s; see Table 2.

### 6.6. Player rankings

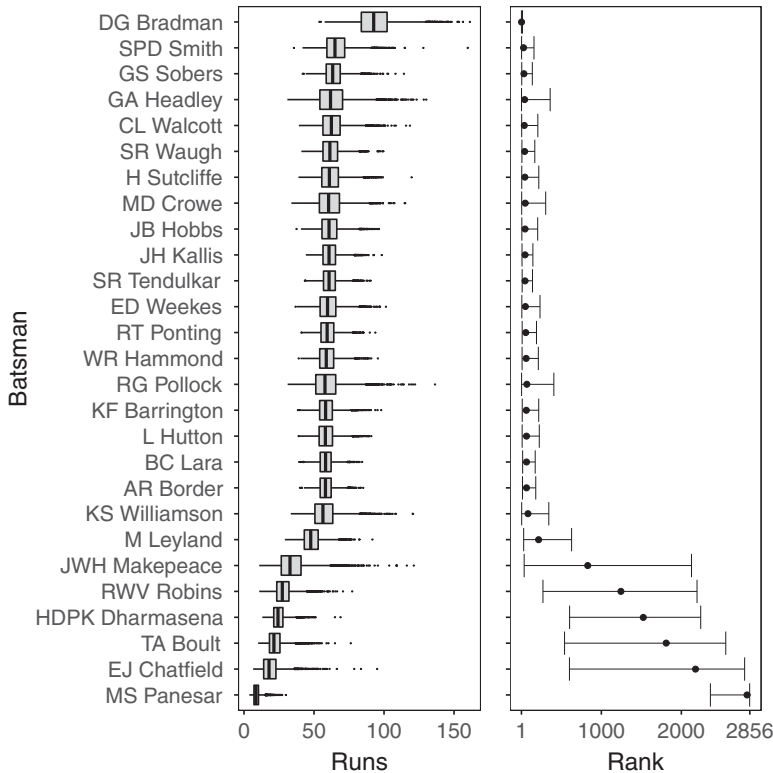
The posterior distributions of mean runs scored by the top 30 ranked players are shown as boxplots in Fig. 6, with numerical summaries in Table 2. Here the players are listed by their posterior mean of  $(1 - \pi) \exp(\theta)$ , i.e. their expected runs scored at their peak age assuming the year of play is 2017 (no year effect) and batting at home in the first innings of a test match against Australia. It is striking just how far Sir Donald Bradman is ahead of the other batsmen, in terms of posterior mass; his extraordinary average is well known to cricket fans and the plot captures this clearly. The posterior distributions of the players ranked from 2 to 30 are largely similar, with considerable overlap. After Bradman it is unclear who is the next ‘best’ batsman. This point is further underlined by the posterior distribution of each player’s rank, calculated across the MCMC samples. Fig. 6 also shows the median posterior rank, together with equitailed 95% confidence intervals. The numerical summaries for each batsman can be found in Table 2. These



**Fig. 5.** Posterior mean and central 95% bands for the aging profile  $(1 - \pi_i) \exp\{\theta_i + f_i(a)\}$  of various batsmen together with their posterior mean-adjusted runs scored: (a) A. D. Nourse; (b) G. A. Headley; (c) H. H. Gibbs; (d) H. M. Amia; (e) J. B. Hobbs; (f) K. S. Williamson; (g) M. D. Crowe; (h) R. T. Ponting; (i) S. P. D. Smith; (j) S. R. Waugh

are summaries of marginal distributions for each player and not, for example, the most probable joint ranking across all players. Therefore it is possible, and happens here, that no batsman has a posterior median rank of, say, 2. However, given the level of variation in runs scored, it does not seem reasonable to rank batsmen by a single-number summary, be it mean score or median rank. Kimber and Hansford (1993) make a similar argument, stating that ‘it is clear that a one-number summary of the distribution of a batsman’s scores is not enough’. Our rank confidence intervals give a much more reasonable measure of rank position and its uncertainty.

The interval for Bradman’s rank is quite narrow, ranging from rank 1 to rank 14. There is little difference in the career batting averages of many players after Bradman and this is borne out in the spread of the confidence intervals for the rankings, which are largely similar and noticeably wide. It is interesting to see the level of (posterior) uncertainty on player rankings. Fig. 6 shows confidence intervals for the top 20 players along with players ranked 100th, 500th and every 500th player thereafter up to the 2500th player and the final player, ranked 2855th. The high level of posterior uncertainty in these ranks chimes with a remark by Goldstein and Spiegelhalter (1996) when comparing institutional performances, that ‘such variability in rankings appears to be an inevitable consequence of attempting to rank individuals with broadly similar performances’. A full list of the ability scores and ranks for all 2855 batsmen can be found via the RShiny application at <https://petephilipson.shinyapps.io/BatsmenRankings/>.



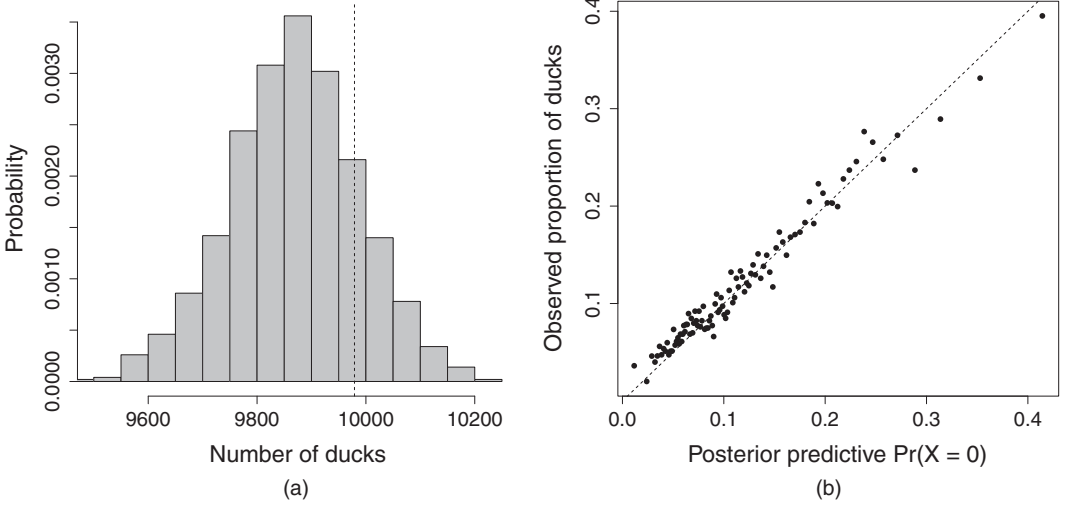
**Fig. 6.** Posterior distributions of some players' abilities  $\exp(\theta_i)$  and their ranking confidence intervals

There are two established rankings lists with which we can compare our rankings. The first is the traditional rankings by career test batting average and the second is the 'Reliance ICC best-ever test championship rating' (ICC) list. These two differ in that the former is a single measure across a player's entire career whereas the latter is the maximum of a dynamic index which places a greater emphasis on recent innings. Our approach could be considered to be a compromise between these two systems. Overall, of the top 30 in our rankings by posterior mean runs scored, we have 23 in common with all-time highest career batting average rankings, and 19 in common with the ICC rankings. Five batsmen in Table 2 do not appear in either of these established ranking lists; these batsmen (with our ranking by posterior mean runs and 95% confidence interval for their rank) are Waugh 6 (7, 168), Crowe 8 (4, 304), Border 19 (14, 180), Williamson 20 (4, 343) and Flower 28 (12, 274). This illustrates a central problem in ranking batsmen by a single-number summary when there is a high level of innings-to-innings variation in runs scored by each batsman. In particular the traditional ranking does not adjust for any covariate information. The ICC rank does adjust for opposition and pitch effects but is empirical and has some other *ad hoc* adjustments. Neither system adjusts for the censoring (not-out innings) problem in a way that takes account of player ability.

### 6.7. Model fitting

We can study the ability of the model to predict ducks (0 scores) by looking at the (model-based) posterior predictive probability of a duck and seeing how this correlates with observed ducks. This predictive probability is calculated by averaging a typical model-based probability





**Fig. 7.** (a) Posterior predictive distribution of total ducks ( $\cdot$ , observed number in the data) and (b) observed proportion of ducks against centiles of posterior predictive probabilities of a duck

$P(X_{ijk} = 0 | \kappa, \eta_i, \pi_i)$  over the uncertainty in the posterior distribution. Therefore we estimate these predictive probabilities by using

$$P(X_{ijk} = 0 | \mathbf{x}, \mathbf{c}, \mathbf{d}) = \frac{1}{N} \sum_{l=1}^N P(X_{ijk} = 0 | \kappa^{(l)}, \eta_i^{(l)}, \pi_i^{(l)}),$$

where  $\{\kappa^{(l)}, \eta_i^{(l)}, \pi_i^{(l)} : l = 1, \dots, N\}$  is the posterior sample. Fig. 7 shows a summary of this information. Fig. 7(a) shows the (posterior) predictive distribution for the total number of ducks in the data set and confirms that this is consistent with its observed data value. In Fig. 7(b), the predictive probabilities have been first grouped into centiles and then the observed proportion of ducks in each centile has been plotted against the mid-point of each centile. The plot shows good agreement between the model predictions and observed proportions as there is little deviation from the 45° line. Fig. 1 in the on-line supplementary materials shows plots that are similar to Fig. 7(b) but gives a more comprehensive picture. Instead of just showing the calibration of duck predictions, this plot contains that for all numbers of runs scored (grouped into intervals, typically of size 10). Overall these plots show that, although the model does not provide a perfect calibration, it does give a fairly accurate description of runs scored.

## 7. Discussion

The data clearly show that there is considerable within-batsmen variability in cricket scores and there is demonstrable evidence that batsmen are especially vulnerable at the beginning of their innings. Also the standard cricket batting average measure makes the unreasonable assumption that run scoring follows a geometric distribution. Further the zero-inflated random-effects Poisson model (with log-linear factors) gives a good description of the runs scored in test matches. In terms of ranking players, we found that Sir Donald Bradman, unsurprisingly, was the best player (under the model) and there was relatively little uncertainty about his ranking. However, there was considerable uncertainty in the rankings of players lower down the list.

We compared our rankings with those of two established lists: one list by career test batting

average and, the other, the 'ICC best-ever test championship rating' list. Not surprisingly we found disagreement between all three lists. This illustrates a central problem in ranking batsmen by a single-number summary when there is a high level of innings-to-innings variation in runs scored by each batsman. In these circumstances it is more appropriate to summarize a career by a distribution which recognizes the uncertainty in these single-number summaries. In this paper we look at the player's overall ability within a model which accounts for the high level of innings-to-innings variation, various cricket-specific factors (not-out innings; zero inflation) and adjusts for various important player-independent factors. Even without such adjustments, simple data averages can easily mislead as some batsmen play relatively few innings: the five-number summary for career innings played is 1–4–12–35–329. We summarize our understanding of the player's ability by a distribution or an interval which accounts for uncertainty. These summaries are impacted less by circumstance and luck (such as when a leg-before-wicket decision goes in the batsman's favour and he makes a big score) than any fundamental difference in ability.

The model represents the quality of the opposition through dynamic opposition-specific parameters to capture potential changes in the performance of test playing countries over time, such as periods of notable strength and weakness. Other factors were considered for inclusion in the model but omitted because of data limitations or in the interests of model parsimony. The effect of playing at home was explicitly accounted for, and this could be subcategorized further into individual test match grounds. However, although some grounds such as Lord's and the Sydney Cricket Ground have been staples on the test match roster, there have been many changes in venues used in the subcontinent and so insufficient data are available to be able to account for individual stadium effects. Test matches are typically played as part of a series but data on the match number within a series were not available in our data set. Similarly, we might take account of the position of the batsmen in the batting order. However, we believe that batting position is chosen to suit the individual characteristics of each player to maximize runs scored, and so we leave out this factor from our model.

An obvious extension of this work would be to apply it to the performance of both batsmen and bowlers. The approach could be further extended to analyse data from 1-day internationals, which, despite being an international sport since only 1971, has already seen around 3900 fixtures take place. This equates to almost the same amount of data as used here for test matches since 1-day internationals feature one innings per team per match. A larger metamodel simultaneously analysing the test and 1-day international batsmen and bowlers could also help in addressing the issue of ranking players as such a model would have the potential to identify not only the quantity of the runs but also to refine attempts to ascertain their 'quality' by explicitly factoring in more granular data relating to the opposition, such as the strength of the bowling attack in a given innings. Twenty20 cricket is another avenue for future work but currently there may be insufficient data for an analysis of the type used in this work.

## Acknowledgements

We thank the Associate Editor and two reviewers for their insightful comments which have improved the paper.

## References

- Baio, G. and Blangiardo, M. (2010) Bayesian hierarchical model for the prediction of football results. *J. Appl. Statist.*, **37**, 253–264.
- Baker, R. D. and McHale, I. G. (2014) A dynamic paired comparisons model: who is the greatest tennis player? *Eur. J. Oper. Res.*, **236**, 677–684.

- Berry, S., Reese, S. and Larkey, P. D. (1999) Bridging different eras in sports. *J. Am. Statist. Ass.*, **94**, 661–686.
- Brewer, B. J. (2008) Getting your eye in: a Bayesian analysis of early dismissals in cricket. *Preprint arXiv:0801.4408*.
- Cameron, A. and Trivedi, P. (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. *J. Appl. Econometr.*, **46**, 347–364.
- Duckworth, F. C. and Lewis, A. J. (1998) A fair method for resetting the target in interrupted one-day cricket matches. *J. Oper. Res. Soc.*, **49**, 220–227.
- Duckworth, F. C. and Lewis, A. J. (2004) A successful operational research intervention in one-day cricket. *J. Oper. Res. Soc.*, **55**, 749–759.
- Durbach, I. N. and Thiart, J. (2007) On a common perception of a random sequence in cricket. *S. Afr. Statist. J.*, **41**, 157–183.
- Elderton, W. (1945) Cricket scores and some skew correlation distributions: (an arithmetical study). *J. R. Statist. Soc.*, **108**, 1–11.
- Gamerman, D. (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statist. Comput.*, **7**, 57–68.
- Gilovich, T., Vallone, R. and Tversky, A. (1985) The hot hand in basketball: on the mis-perception of random sequences. *Cogn. Psychol.*, **17**, 295–314.
- Goldstein, H. and Spiegelhalter, D. J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). *J. R. Statist. Soc. A*, **159**, 385–443.
- Greene, W. (2008) Functional forms for the negative binomial model for count data. *Econ. Lett.*, **99**, 585–590.
- Gurmu, S. (1997) Semi-parametric estimation of hurdle regression models with an application to medicaid utilization. *J. Appl. Econometr.*, **12**, 225–242.
- Gurmu, S. (1998) Generalized hurdle count data regression models. *Econ. Lett.*, **58**, 263–268.
- Hougaard, P., Lee, M. L. and Whitmore, G. A. (1997) Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, **53**, 1225–1238.
- Jones, M. (2007) Home advantage in the NBA as a game-long process. *J. Quant. Anal. Sports*, **3**, no. 4, 1–14.
- Kimber, A. C. and Hansford, A. R. (1993) A statistical analysis of batting in cricket. *J. R. Statist. Soc. A*, **156**, 443–455.
- Lambert, D. (1992) Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.
- Marchi, M. and Albert, J. (2013) *Analyzing Baseball Data with R*. Boca Raton: Chapman and Hall–CRC.
- McHale, I. G. and Asif, M. (2013) A modified Duckworth–Lewis method for adjusting targets in interrupted limited overs cricket. *Eur. J. Oper. Res.*, **225**, 353–362.
- Plummer, M. (2004) JAGS: just another Gibbs sampler. (Available from <https://sourceforge.net/projects/mcmc-jags/>.)
- Plummer, M., Best, N., Cowles, K. and Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.
- Pollard, R., Benjamin, B. and Reep, C. (1977) Sport and the negative binomial distribution. In *Optimal Strategies in Sports* (eds S. P. Ladany and R. E. Machol), vol. 5, pp. 188–195. Amsterdam: North-Holland.
- Pollard, R. and Pollard, G. (2005) Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *J. Sports Sci.*, **23**, 337–350.
- Radicchi, F. (2011) Who is the best player ever?: a complex network analysis of the history of professional tennis. *PLOS One*, **6**, no. 2, article e17249.
- R Core Team (2014) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ridout, M., Hinde, J. and Demtrio, C. G. (2001) A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**, 219–223.
- Rohde, N. (2011) An economical ranking of batters in Test cricket. *Econ. Pap.*, **30**, 455–465.
- Scarf, P., Shi, X. and Akhtar, S. (2011) On the distribution of runs scored and batting strategy in test cricket. *J. R. Statist. Soc. A*, **174**, 471–497.
- Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M. and Meze-Hauksen, E. (2013) A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *J. Appl. Statist.*, **62**, 85–100.
- Silva, R. M., Manage, A. B. and Swartz, T. B. (2015) A study of the powerplay in one-day cricket. *Eur. J. Oper. Res.*, **244**, 931–938.
- Stern, S. (2009) An adjusted Duckworth–Lewis target in shortened limited overs cricket matches. *J. Oper. Res. Soc.*, **60**, 236–251.
- Tversky, A. and Gilovich, T. (1989) The ‘hot hand’: statistical reality or cognitive illusion? *Chance*, **2**, no. 4, 31–34.
- Wood, G. H. (1945) Cricket scores and geometrical progression. *J. R. Statist. Soc.*, **108**, 12–39.

#### Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplementary materials for On the ranking of Test match batsmen’.